

Andrew Schechtman-Rook

rook166@gmail.com
(917)-836-4267

<https://github.com/AndrewRook>

TECHNICAL SKILLS

Programming: Python (numpy, scipy, pandas, sklearn, xgboost, matplotlib), shell scripting

Model Development & Deployment: GBMs, random forests, linear & logistic regression, nonlinear optimization, parallel & distributed computing, containerization, APIs

Databases, Orchestration, Web Design: Flask, MySQL/PostgreSQL, Airflow, Prefect, Snowflake

Cloud Computing/Devops: AWS, CircleCI, Jenkins, GitHub Actions

RELEVANT PROJECTS

Ptplot: <https://github.com/AndrewRook/ptplot>

- Python package for quickly creating interactive visualizations of player tracking data.
- Visualizations can be viewed in notebooks or embedded in websites.

NFLDash: <https://andrewrook.github.io/NFLDash>

- Web dashboard allowing users to interactively filter events on 15+ dimensions, then export results to disk.
- Built in HTML and Javascript with d3.js and crossfilter.js.

NFLWin: <https://github.com/AndrewRook/NFLWin>

- Python package for calculating Win Probability of NFL data.
- One of the first fully open (code & algorithm) published Win Probability models.

WORK EXPERIENCE

Director, Data Science: Capital One

2023-Present

Technical lead on Card Credit Innovation team, building and maintaining core credit infrastructure

- Designed and built a Python package to streamline core business metrics SQL calculations, used by over 50 analysts for critical reporting needs.
- Prototyped an LLM-based approach to automating SQL query generation using RAG, with accuracies up to 80% on test datasets.
- Mentored multiple data scientists, including starting an individual-contributor focused talk series.

Senior Manager, Data Science: Capital One

2018-2023

Model developer and technical lead for Upmarket Card Data Science and core developer on valuation model infrastructure team

- Guided ongoing development of the core credit card valuations model scoring platform, delivering regular releases of new and updated models while improving the robustness and maintainability of platform infrastructure.
- Led technical development of model monitoring tools, mentoring three junior data scientists to deliver a maintainable package on time and to spec.
- Deployed the first cloud-based credit card underwriting model in the company via a dockerized Python API, with an estimated incremental value of \$35MM per year.

Manager, Data Science: Capital One

2016-2018

Underwriting model deployment and tooling subject matter expert

- Led development of a prototype language-agnostic automated machine learning model deployment framework for cloud-based applications, influencing the development direction for the company-wide credit card application processing platform.
- Created the longest-lived, most successful internal data science tool in the company, used in production models by dozens of data scientists across multiple lines of business.

Principal Data Scientist: Capital One Labs

2014-2016

Core performer and project manager for an internal data science training program

- Implemented a novel approach to deliver internal technical trainings, providing over 5000 hours of classes with no instructors.
- Programmed and deployed an interactive course completion dashboard using Flask and dc.js to provide progress reports to individual students as well as company leadership.

Research Associate: University of Wisconsin-Madison

2014

Postdoctoral researcher

- Devised metrics to improve correspondence between numerical models and astronomical data. Implemented in highly optimized Python, was able to refine agreement by up to 20% with minimal increase in computation time.
- Built a fast Voronoi Tessellation algorithm to adaptively bin images, preserving spatial resolution while maximizing signal in images with over one million pixels.
- Trained and mentored undergraduate and graduate students in programming, data analysis and statistical methods.

Research Assistant: University of Wisconsin-Madison

2007-2013

Graduate student

- Developed non-linear Levenberg-Marquardt χ^2 fitting algorithms using a combination of Python and C++ to constrain models of spiral galaxies to data.
- Employed on-campus distributed computing resources to perform large-scale modeling in parallel, using over 20 years of computer time in 1 month.
- Assembled a hybrid C++/Python processing pipeline to clean, register, and mosaic thousands of high-resolution images with minimal user intervention, resulting in a factor of 10+ increase in analysis precision.
- Created a genetic algorithm in C++ to efficiently fit galaxy models with unusually large numbers of free parameters to high-resolution images.

EDUCATION

PhD, Astronomy, University of Wisconsin-Madison

December 2013

MS, Astronomy, University of Wisconsin-Madison

June 2009

BS, Astronomy, Case Western Reserve University

May 2007